Article type: Original article

Environmental controls on the global distribution of shallow-water coral reefs

Authors:

E. Couce [1,2], A. Ridgwell [2], E.J. Hendy [1,3]*

[1] *School of Earth Sciences, Queens Road, University of Bristol, Bristol BS8 1RJ, UK*

[2] *School of Geographical Sciences, University of Bristol, Bristol BS8 1SS, UK*

[3] *School of Biological Sciences, Woodland Road, University of Bristol, Bristol BS8 1UG, UK*

*Corresponding author: e.hendy@bristol.ac.uk, Tel: +44 (0)117-331-5003

Running head: global distribution models of shallow-water coral reefs

Word count (abstract): 349 (main text body): 5818

ABSTRACT

Aim

Elucidating the environmental limits of coral reefs is central to projecting future climate change impacts on these ecosystems and their global distribution. Recent developments in Species Distribution Modelling (SDM) and the availability of comprehensive global environmental datasets provide an opportunity to reassess the environmental factors that control coral reef distribution at the global scale as well as compare the performance of different SDM techniques.

Location

Shallow waters worldwide.

Methods

The SDM methods used include a presence-only technique: Maximum Entropy (MaxEnt), and two presence/absence methods: Classification and Regression Trees (CART) and Boosted Regression Trees (BRT). The predictive variables considered include sea surface temperature (SST), salinity, aragonite saturation state (Ω_{Arag}), nutrients, irradiance, water transparency, dust, and cyclone frequency and intensity. For many variables both mean and SD were considered, and at weekly, monthly and annually-averaged timescales. All were transformed to a global 1°x1° grid to generate coral reef probability maps for comparison with known locations in the ReefBase (2000) dataset.

Results

Model performance was compared in terms of Receiver-Operating-Characteristic (ROC) curves and Area-Under-the-Curve (AUC) scores. BRT consistently outperformed other

methods (AUC 0.90), although MaxEnt was also within acceptable performance limits (AUC 0.88). The dominant environmental predictors were the temperature variables; annual mean SST, and monthly and weekly minimum SST, followed by and with their relative importance differing between regions; nutrients, light availability and aragonite saturation. Potential geographical bias was explored via misclassification maps of false positive and negative errors on test data. While no systematic bias was found between major coral provinces, false negatives were more likely for cells that corresponded to 'marginal' non-reef-forming coral communities e.g. southern Queensland, Australia and Bermuda.

Main conclusions

Although SST-related variables dominate the coral reef distribution models, contributions from nutrient concentration, aragonite saturation and light availability were critical in developing models of reef presence in regions such as the Bahamas, south Pacific and Coral Triangle (particularly Sulu Sea and Spratly Is.). Agreement between BRT and MaxEnt models gives predictive confidence that they can be used to explore the environmental limits of coral reef ecosystems at a spatial scale relevant to global climate models (~1°x1°).

Keywords

Boosted Regression Trees (BRT), Classification and Regression Trees (CART), coral reef biogeography, environmental limits, Maximum Entropy (MaxEnt), Species Distribution Modelling (SDM).

INTRODUCTION

Coral reefs cover less than 0.1% of the world's oceans (Spalding & Grenfell, 1997) but constitute one of the most biologically diverse and economically valuable ecosystems on Earth (e.g. Reaka-Kudla, 1997; Moberg & Folke, 1999). These 'oases' support levels of productivity several hundred times higher than that of surrounding generally nutrient-poor areas (Hatcher, 1988). Symbiosis between scleractinian 'stony' coral and photosynthesising unicellular algae (*Symbiodinium spp.*) make this possible, but also limit coral reefs to the warm, shallow and clear waters of the tropics where sunlight is readily available.

Reefs are vulnerable to anthropogenic disturbance, including overfishing and overharvesting of coral, physical damage from destructive fishing and shipping, increased sedimentation due to land use changes, and reduced water quality resulting from coastal runoff. Combinations of these stresses are highly likely responsible for observed declines of these emblematic ecosystems (reviewed in Carpenter *et al.*, 2008). Coral reefs now face additional pressures exerted by climate change. The increase in global mean Sea Surface Temperature (SST) of at least 0.5°C since 1850 (Rayner *et al.*, 2006) has been implicated in widespread coral bleaching events, and hence increased mortality (Carpenter *et al.*, 2008). At the same time, the ocean uptake of CO_2 has decreased seawater pH by an estimated -0.1 units since 1750 (Bindoff *et al.*, 2007). Ocean acidification may compromise the ability of reef-building corals and other carbonate-producing organisms to calcify rapidly enough to maintain reef structures (Kleypas *et al.*, 2006). Recent observations of reduced calcification rates in massive

Porites colonies from the Great Barrier Reef (GBR) are consistent with this already occurring (Cooper *et al.*, 2008; De'ath *et al.*, 2009), and in the near future, calcification could potentially become marginal or strongly inhibited in most areas where reefs are found today (Hoegh-Guldberg *et al.*, 2007; Cao & Caldeira, 2008).

Central to projecting future impacts of climate change and ocean acidification on these ecosystems is a better understanding of what controls the biogeography of coral reefs today. This question has been pursued since Dana (1843), but the seminal study of global reef distribution as a function of environmental factors is Kleypas *et al*. (1999), in which lower limits of SST, aragonite saturation state (Ω_{Arag}) and light availability were identified as key factors determining coral reef distribution. The range of Ω_{Arag} values within which current coral reefs were found has proved particularly influential, and has been adopted as a critical threshold on coral reef distribution under future climate change and ocean acidification scenarios (e.g., Hoegh-Guldberg *et al.*, 2007; Cao & Caldeira, 2008).

Since the 1999 Kleypas *et al*. study, new global ocean environmental datasets have become available and significant advances have been made in the emerging field of Species Distribution Modelling (SDM, see Guisan & Zimmermann, 2000; Elith *et al.*, 2006; and Franklin, 2009 for an overview). These new statistical methods have been developed to meet an increasing demand from conservation planning, ecological and risk assessments, and ecosystem management amongst other fields. Assessment tools to evaluate and compare models' performance in different contexts have been

92 developed and improved (Fielding & Bell, 1997; Manel *et al.*, 2001; Johnson &
93 Gillingham, 2005), and the limitations of SDM better assessed and documented. The
94 latter include the effects of sample size (Hernandez *et al.*, 2006) and sample bias (Costa
95 *et al.*, 2010), spatial scale (Seo *et al.*, 2009), species prevalence (Manel *et al.*, 2001),
96 variations amongst predictions from different models (Johnson & Gillingham, 2005),
97 and limitations of their use in global change studies (Thuiller, 2004; Botkin *et al.*, 2007).
98 SDM methods have been used to analyse environmental controls on coral reef
99 distribution on a regional scale, for example cross-shelf and latitudinal changes of water
100 quality measures within the GBR (De'ath & Fabricius, 2010), and deep water
101 azooxanthellate coral on a global scale (Tittensor *et al.*, 2009).
102
103 Here we make use of recent methodological developments to revisit the question of
104 what environmental factors control surface water coral reef distribution at the global
105 scale. The predictive variables we have considered that may directly influence (or co-
106 vary with other driving factors) coral reef distributions, include: temperature, aragonite
107 saturation state, salinity, nutrients, indicators of upwelling (e.g. dissolved oxygen and
108 nutrient concentrations below the photic zone), irradiance, water transparency, dust,
109 and cyclone frequency and intensity. For many of these variables we considered both
110 mean and SD at weekly, monthly, and annual time periods. We compare the predictions
111 and performance of three different SDM statistical methods: Maximum Entropy
112 (MaxEnt), Classification And Regression Trees (CART) and Boosted Regression Trees
113 (BRT). The dominant environmental controls in different geographical areas were
114 identified by studying a series of different variable combinations and mapping the

predicted coral reef distribution. Although too coarse for elucidating detailed biological requirements of coral organisms, the 1°x1° spatial scale we use in our analysis is compatible with current global climate models. Combining climate model output with these statistical environmental models of potential reef distribution could provide a valuable tool for projecting future reef distributions under anthropogenic global warming and ocean acidification.

MATERIALS AND METHODS

Reef location data

The point coordinates of over 11,000 coral reefs used in the study were obtained from 'ReefBase', version 2000 (ReefBase, 2000). ReefBase is a composite of published and unpublished sources (e.g. peer-reviewed journal articles, field surveys, and other reports) of coral reef presence. The database includes entries for both 'reef' and 'non-reef coral communities'; the latter characterised by the inability to accumulate calcium carbonate. The presence records were mapped onto a 1°x1° global grid, with the cells designated as 'presence' sites if they contained one or more 'ReefBase' entries (a total of 1242 presence sites), and 'absence' sites if the grid cell did not contain any ReefBase entries. A fixed-degree projection was used instead of equal-area for convenience because most environmental datasets are provided on degree-based grids. A similar global analysis for stony corals on seamounts by Tittensor et al. (2009) found no difference between using 1°x1° resolution and an equal-area projection.

Environmental data

The environmental variables tested in our analysis are listed in Table 1. This includes variables previously considered by Kleypas *et al*. (1999), plus nitrate and phosphate concentrations at 100m water depth, dissolved oxygen concentrations on the surface and 100m water depth, current strength, cyclone activity, atmospheric dust content, and SST variability at various time scales. Of these, nutrient and dissolved oxygen concentrations at 100m water depth were tested as proxies for upwelling regions. Surface dissolved oxygen provides an additional non-linear response to temperature via its solubility. Current strength and cyclone action were considered as proxies for potential physical destruction of structural framework and inhibition of carbonate accumulation (Riegl, 2001), while high levels of dust have previously been linked to reef decline and reduced coral settlement (e.g. Shinn *et al.*, 2000). Finally, weekly, monthly, and inter-annual standard deviation (SD) of SST captures the degree to which a site is exposed to highly variable conditions. For example, SST fluctuations can be associated with stratification and high UV penetration, or climate phenomena such as the El Niño Southern Oscillation, all contributing factors to recent mass bleaching events (e.g. Glynn & Colgan, 1992; Fitt *et al.*, 2001; Berkelmans *et al.*, 2004).

As summarised in Table 1, mean annual and monthly values of SST (Locarnini *et al.*, 2006), and salinity (Antonov *et al.*, 2006), together with the annual averages of nitrate, phosphate (Garcia *et al.*, 2006a) and dissolved oxygen concentrations (Garcia *et al.*, 2006b) were obtained from the World Ocean Atlas (WOA) 2005 database. Satellite-derived daily irradiance (1991-1993) is from the ISCCP project (Bishop & Rossow,

1991; Bishop *et al.*, 1997) while depth of light penetration was estimated following Kleypas *et al.* (1999), using light irradiance and satellite measurements of the diffuse attenuation coefficient of (490 nm) light from GlobColour (2008). Aragonite saturation (Ω_{Arag}) data was calculated using 'CO2sys' version 1.05 (van Heuven *et al.*, 2009) from GLODAP (Key *et al.*, 2004) alkalinity and total CO_2 together with WOA (2005) temperature, salinity, silicate and phosphate and the carbonate dissociation constants of Dickson and Millero (1987) (Mehrbach et al's 1973 refit, see CO2sys' documentation for details). Areas missing GLODAP data (e.g. Caribbean) were given Ω_{Arag} values taken from the UVic Earth system model (Weaver *et al.*, 2001; Turley *et al.*, 2010) following the methodology of Cao and Caldeira (2008). Global dust data was obtained from NCAR CCSM3 model output (Mahowald *et al.*, 2006). The intensity of cyclone activity was the "sum of wind" product obtained from UNEP/GRID-Europe (2009) and based on data of >2800 historical cyclone events and modelled wind speed for 1977-2006. Weekly SST data (1982-2009) was from NCEP Reynolds Optimally Interpolated SST product (Reynolds *et al.*, 2002) and was used to estimate the SD of average monthly SST in both January and July and to obtain average maximum and minimum weekly temperatures. Finally, annual average, monthly maximum and minimum current speeds (2002-2006) were calculated from SODA version 2.0.4 (Carton & Giese, 2008).

Grid and Mask

All environmental variables were implemented on a 1°x1° grid between -60° and 60° latitude. A shallow-water mask (Fig. 1) was created from 30 second resolution SRTM30 Plus bathymetry data (Becker *et al.*, 2009). The ±60° latitude study region was

183 restricted to areas with a benthic substrate depth that solar light could reach, as
184 computed from the amount of Photosynthetically Active Radiation (PAR) reaching the
185 surface and the water's attenuation coefficient at a wavelength of 490 nm. To minimise
186 the number of isolated sea-mounts not resolved due to limitations in the resolution of
187 the bathymetry data, the mask was established as the grid cells with regions shallow
188 enough to be within twice the mean annual depth of light penetration. A total of 4649
189 grid cells passed the mask and contained 1115 of the 1242 designated reef presence
190 sites (see Fig. 1). The remaining 127 presence cells falling outside the mask were
191 discarded (>90% plotted on land or were incorrect/incomplete entries in the ReefBase
192 dataset). Variables missing information on a reef cell were approximated by linear
193 interpolation of neighbouring cells.
194
195 **SDM methods**
196 Three statistical methods were used to determine the relationship between the global
197 distribution of reefs and the environmental variables studied. These were Maximum
198 Entropy Modelling (MaxEnt), Classification and Regression Tree (CART), and Boosted
199 Regression Trees (BRT).
200
201 MaxEnt (Phillips *et al.*, 2006; Phillips & Dudik, 2008) is a widely used (e.g. Tittensor *et*
202 *al.*, 2009; Bradley, 2010; Ficetola *et al.*, 2010) presence-only technique for the
203 prediction of species geographic distributions (i.e. predictions are based only on the
204 environmental conditions of sites of known occurrence). The method assumes that
205 environmental factors act as constraints on the distribution of a species, and that within

206 those constraints, the species will tend to occupy all available habitat in a way that
207 maximises entropy (i.e., the species distribution is not restricted by any additional
208 unknown constraints). We used MaxEnt version 3.3.2, with the convergence threshold
209 value set to 10^{-5} and a maximum number of iterations of 500. The output was obtained
210 as the logistic prediction for presence of reefs, with values ranging from 0 (unsuitable)
211 to 1 (optimal conditions). MaxEnt's logistic prediction can be interpreted as the
212 probability of finding the species for a sampling effort similar to the one that produced
213 the training dataset (see Phillips & Dudik, 2008). As the sampling effort is typically
214 unknown, MaxEnt defaults to assuming an average 50% probability in locations where
215 conditions are favourable.
216
217 CART (Breiman *et al.*, 1984; De'ath & Fabricius, 2000) and BRT (Friedman, 2001;
218 De'ath, 2007) are decision-tree based statistical techniques. A single tree (CART) is built
219 by repeatedly splitting the data using a simple rule each time, so that the homogeneity
220 of the resulting groups is maximised. The optimal size of the tree can be chosen from
221 different stopping criteria, such as cross-validation, to avoid over-fitting to training data
222 (De'ath & Fabricius, 2000). Decision trees have the advantages of being able to handle
223 non-linear relationships, high-order interactions, categorical or numerical data, and
224 missing data. For the BRT method a series of trees is produced instead of a single one.
225 Each tree is grown on reweighted versions of the original data, with an ever-increasing
226 weight being assigned to the cases misclassified by previous trees. The final prediction
227 is obtained by the weighted average of predicted values across the sequence of trees.

228 BRT models have been shown to be extremely good predictors in ecological studies
229 (Leathwick *et al.*, 2006; Moisen *et al.*, 2006).
230
231 Both CART and BRT were fitted in R (version 2.10.1, R Development Core Team, 2009).
232 The spatial data was processed using the *sp* library, version 0.9-65 (Pebesma & Bivand,
233 2005; Bivand *et al.*, 2008), while single trees were generated with version 3.1-46 of the
234 *rpart* library (Therneau & Atkinson, 2010) and boosted trees with version 1.6-3.1 of the
235 *gbm* library (Ridgeway, 2007). The method adopted to generate boosted trees is the one
236 proposed by Elith *et al.* (2008). Generating a BRT model requires specifying a number of
237 parameters that control the process. These include the Tree Complexity (TC), which
238 determines the size of each individual tree, the Learning Rate (LR), which establishes
239 each tree's contribution to the final prediction, and the Bagging Fraction (BF), or
240 percentage of random test data used to generate each of the trees. For this study, in
241 order to generate each BRT model we tested TC values of 1, 3, 5, 7, and 10, LR values of
242 0.5, 0.1, 0.05, 0.01, 0.005, and 0.001, and BF of 0.5, 0.6 and 0.75 (values suggested by
243 Elith *et al.*, 2008). The combination that optimised the model's predictive deviance from
244 coral reef observations within reasonable time constraints was chosen in each case. An
245 example of the procedure is given in Appendix S1.1 in Supporting Information.
246
247 **Methods for model comparison**
248 The output from the three models was compared in terms of Receiver Operating
249 Characteristic (ROC) curves (reviewed in Zweig & Campbell, 1993); graphical plots of
250 the true positives versus false positives for a binary classifier system. A curve in true-

positives vs. false-positive space is obtained as the cut-off value for the probability at which a cell is considered a presence site (the 'discrimination threshold') is allowed to change. The ROC method, in addition to being independent of the discrimination threshold, is less biased by the prevalence of the species under study than other methods commonly used to test SDM output (Manel *et al.*, 2001). Although ROC was originally developed for presence/absence techniques, it can also be applied to presence-only models. Typically this involves comparing prediction at presence sites with prediction at sites chosen randomly from the study region (called 'pseudo-absences'). In our case, as absence data was available and used in the development of the presence/absence models, in order to establish a meaningful comparison, MaxEnt's ROC curve was also generated using true absence data. For techniques based on decision trees (CART and BRT), model performance is indicated by a single point in ROC-space, corresponding to a 50% probability of presence of the species/ecosystem being studied. While the discrimination threshold has an obvious cut-off, the rest of the ROC curve can still be obtained if the value is allowed to change. Average Area Under the Curve (AUC) scores (Bradley, 1997) were obtained from the output of 10 different model runs, generated each time from a random choice of 75% of the total data (training data) and tested on the remaining 25% (evaluation data).

Methods for assessing environmental variables as model predictors

Some of the 31 environmental variables used in this study (Table 1) are strongly correlated (Appendix S3; Fig. S3.1). In order to explore whether all potential environmental variables were necessary to the overall prediction and assess how the

importance of different variables differ geographically, we created an ensemble of models using different combinations of environmental variables (Table 1). In Appendices S1 and S2 we included the results of a number of standard tests for exploring SDM output, such as 'variables' contribution indices' (Tables S1.1 and S2.1), which aim to measure the relative relevance of each variable to the model, and 'marginal response curves' (Figs. S1.6 S2.2, S2.3), which are graphical representation of the effect on the model's output of variations of a single variable while keeping the rest fixed at their average value.

RESULTS

Prediction maps

The results from all three SDM techniques (Fig. 2) are consistent with the observed major regions of coral reef presence (Fig 1). The probability maps (Fig. 2) use the *OPTIMAL* set of variables (Table 1), identified as the optimal combination of environmental variables to study present coral distribution (results from other models will be discussed below). Modelled coral reef probabilities are plotted for the shallow water mask with values from 0 (low probability; blue) to 1 (high probability; red). MaxEnt (Fig. 2b) differs from the BRT model (Fig. 2a) in having a compressed probability distribution, with predicted coral reef regions predominantly assigned logistic predictions close to 0.5 with limited high-probability sites. This is a characteristic of the MaxEnt method's default logistic output, which assumes that typical presence localities have an average probability of presence of about 0.5. Tree-based methods, in contrast, attempt to assign high values to presence sites, and their

297 predictions are not directly comparable to MaxEnt's. In addition, the CART approach
298 demonstrates noticeable discretisation as a result of the small number of final states (or
299 terminal nodes) available (Fig 2c). Enlargements of the BRT and MaxEnt prediction
300 maps for the Coral Triangle region are given in Supporting Information (Fig. S1.9 &
301 S2.7).
302
303 **Histograms and misclassification maps**
304 The predicted probability values for each model can also be visualised as histograms
305 (Fig. 3). These histograms separate the presence and absence sites according to
306 ReefBase data, and were obtained from 30 model runs, each using a randomly chosen
307 75% of the data for training, and the remaining 25% for model evaluation. The values
308 assigned to each cell correspond to the average of all runs in which the cell was selected
309 to evaluate the model. Fig. 3 also shows the percentage of false positives (cells without
310 reefs where the model predicts presence) and false negatives (cells with reefs where
311 the model predicts absence). BRT correctly predicted coral presence in over 60% of
312 actual reef cells, a percentage that dropped to 50% for CART. MaxEnt distribution has a
313 clear peak at 0.5, with 65% of reef cells being assigned logistic prediction values
314 between 0.4 and 0.6. Prediction errors (false positives and negatives) have also been
315 mapped in Fig. 4, to investigate possible spatial bias.
316
317 **Assessment of model performance**
318 The performance of the three SDM methods was assessed by means of Receiver
319 Operating Characteristic (ROC) curves. The curves in Fig. 5 correspond to the *OPTIMAL*

variable set (see Table 1 for variable choice). Ten models were developed, each using a randomly chosen 75% of all the available data for model training; the curves (Fig. 5) correspond to each model's performance evaluation on the remaining 25% data. The average AUC-scores obtained from these curves are shown in Table 2. The presence/absence BRT model was the most successful, with an AUC score of 0.9, while the presence-only MaxEnt model performed almost as well (Fig. 5, Table 2). AUC scores above 0.9 correspond to 'highly accurate' models according to Swets (1988), with the 0.7-0.9 range corresponding to models with some useful applications and scores below 0.7 indicating 'low accuracy. BRT represents an upgrade of CART and clearly outperforms it.

Prediction maps of coral reef distribution were developed with different subsets of the predictive variables using both BRT and MaxEnt (Fig. 6). Table 1 lists the model names and the selected variables used to develop each of the models. The probability histograms for each model and the comparison of their performance in ROC space are also provided in Appendix S3 (Figs. S3.2 and S3.3). Despite very similar, often overlapping results in ROC space, the models developed from different variable combinations display distinct spatial distribution predictions of coral reefs presence.

Discussion

Coral biogeography and the spatial structure of environmental control

All SDM models indicate that coral reef distribution is primarily limited to regions of high SST, with high light availability, salinity, aragonite saturation state, and low

343 nutrient concentrations being of secondary significance. Levels of dust, current speed,
344 cyclone intensities, and inter-annual SST variability, are all typically assigned to be of
345 minor importance. The mean percentage contribution values and SD in 10 model runs
346 are given for both BRT and MaxEnt outputs (Tables S1.1 and S2.1). The BRT estimation
347 of percentage contribution takes into account the number of times each variables is
348 selected for splitting weighted by the square improvement of the model as a result of
349 those split and averaged across the sequence of trees (Friedman & Meulman, 2003).
350 Whereas, MaxEnt keeps track of the increase in gain due to each of the variables used in
351 fitting the model, and the normalised percentage contribution of all variables are
352 considered a measure of each variable's relevance to the final result (Phillips, 2006).
353 However, because some of the variables are highly correlated, caution is required in the
354 interpretation of these contribution values. An alternative for MaxEnt is Jackknife tests
355 (Appendix S2; Figs. S2.4-6). In order to explore the spatial structure of environmental
356 controls on coral reef biogeography, the relative contribution of each variable can also
357 be analysed in conjunction with the probability maps (Figs. 6 & 7). Testing different
358 combinations of potential predictive environmental variables provides insight into the
359 control of these highly correlated variables on model response and global distribution
360 of coral reefs.
361
362 In combination, SST variables contribute close to 50%-60% of the environmental
363 controls on coral distributions according to all SDM methods (using *OPTIMAL* model
364 variables; output Fig. 3). This is not surprising. First described in the 19th century (Dana,
365 1843), the high correlation between elevated surface temperature and coral

366 distribution has been well documented (e.g. Stoddart, 1969; Rosen, 1984; Veron &
367 Minchin, 1992). The majority of coral reefs are found in water temperatures with an
368 annual minimum SST greater than 18°C and average minimum weekly temperature
369 above 16°C (reviewed in Kleypas, 2007). The marginal response curves for both BRT
370 and MaxEnt methods indicate threshold values for minimum monthly and weekly SST
371 that are consistent with these estimates (S2 and S3). Mechanisms through which high
372 SST favours the presence of coral reefs include positive physiological effects on growth
373 and calcification rates (Harriott, 1999; Lough & Barnes, 2000), increased reproductive
374 viability (reviewed in Harriott & Banks, 2002), and ecological advantages in the
375 competition with macro-algae (Johannes *et al.*, 1983) linked to nutrient availability and
376 herbivory pressures (McCook *et al.*, 2001; Smith, 2008).
377
378 Models based on SST variables alone (Table 1 model *SST*), however, provide limited
379 explanation of the global distribution of coral reefs (Figs. 6a & 7a) because they fail to
380 account for reefs in higher latitudes, such as the Southern GBR, New Caledonia, South
381 French Polynesia, and Northern Red Sea, South Japan and Ryukyu Islands. The
382 predictions for certain high-temperature regions (e.g. South China Sea/Sulu Sea and
383 South East Papua New Guinea) were also poor. Identifying what environmental factors
384 in addition to SST control distribution at these sites is particularly critical for
385 understanding how reefs will respond under future global change scenarios. Modelled
386 reef distribution is improved by the inclusion of light availability (Figs. 6d & 7d; model
387 *SST_LDepth*), nutrients (Figs. 6e & 7e; model *SST_Nutr*), aragonite saturation (Figs. 6b &
388 7b; model *SST_Arag*), and secondary SST variables (Figs. 6f & 7f; model *SST_2ndSST*).

389 Each model tends to reinforce *SST* predictions due to synergistic interactions, but also
390 demonstrates critical addition information that is needed for correctly modelling
391 particular geographical locations.
392
393 Water clarity, together with solar PAR irradiance, determine the amount of light
394 available and hence limit the depth of reef formation via depth-related control of coral
395 primary productivity and calcification rates (Gattuso *et al.*, 2006). Light availability, and
396 therefore the maximum depth of reef formation, decreases with increasing latitude
397 (Gattuso *et al.*, 2006). However, the addition of light penetration depth variables
398 improves predictions in both the Pacific and Caribbean regions (Figs. 6d & 7d) by
399 extending the latitudinal range of predicted reef habitats (e.g. to include southern
400 islands of French Polynesia). Including light availability also improves predictions for
401 the South China Sea/ Sulu Sea region and the Bahamas.
402
403 In general terms the inclusion of seawater nutrient (phosphate and nitrate)
404 concentrations is expected to exclude modelled reefs from areas enriched by riverine
405 discharge, upwelling along eastern equatorial ocean boundaries and higher latitudes in
406 response to the latitudinal gradient of increasing nutrients. Coral reef manipulation
407 studies reveal a more complex relationship, which is often species and region-
408 dependant and with responses that are typically subtle (e.g. Koop *et al.*, 2001; Dizon &
409 Yap, 2005; Smith *et al.*, 2005). Instead, macroalgae abundance, which generally
410 correlates with nutrient concentrations, may play a decisive controlling role (e.g.
411 Johannes *et al.*, 1983; Birkeland, 1988; De'ath & Fabricius, 2010). Macroalgae have been

412 shown to directly interfere with coral recruitment, suppress coral growth and fecundity,
413 increase disease-related mortality, and compete with coral for space (reviewed in
414 McCook *et al.*, 2001; Smith *et al.*, 2006; Mumby & Steneck, 2008). The inclusion of
415 surface nutrient concentrations is critical for modelling sites in the Southern GBR and
416 Southern Japan (Figs. 6e & 7e). These parameters also have the strongest influence on
417 improving SDM predictions across the Caribbean.
418
419 Aragonite saturation has the largest effect in both BRT and MaxEnt after SST, light and
420 nutrient variables (Figs. 6b & 7b; Appendix Tables S1.1 & S2.1). There is a well-
421 documented relationship between low Ω_{Arag} and reduced coral calcification rates
422 (experimental studies reviewed in Kleypas *et al.*, 2006; e.g. field studies Cooper *et al.*,
423 2008; De'ath *et al.*, 2009). Low cementation of reef framework and high bioerosion
424 rates are also associated with low Ω_{Arag} (e.g. Manzello *et al.*, 2008; Silverman *et al.*,
425 2009). In addition, high magnesium calcite precipitating coralline algae, which play a
426 key role in the cementation of coral rubble into solid reef, are particularly sensitive to
427 carbonate saturation state (Kuffner *et al.*, 2008). The *SST_Arag* model reinforces
428 predictions of reef presence in areas already picked up by the *SST* model, as expected
429 since both datasets are highly correlated (Appendix S3; Fig. S3.1). Although the addition
430 of Ω_{Arag} dramatically improves predictions in areas such as the Spratly Islands / Sulu
431 Sea and Southern French Polynesia (Figs. 6b & 7b), predictions for the Northern GBR
432 deteriorate in the BRT model. The latter result potentially highlights an issue with the
433 quality of the Ω_{Arag} dataset for this region.
434

435 Modelling probability of reef presence in the Southern GBR proves particularly
436 challenging (Figs. 6 &7), although the *OPTIMAL* variable selection is capable of
437 providing the information needed (Fig. 2). Reef formation in this area is apparently
438 controlled by a different combination of conditions than those acting on the Northern
439 section. Of all the variable models presented in Figs. 6 and 7, only the inclusion of
440 nutrients and secondary SST variables improve predictions for the Southern GBR. Other
441 regions challenging to model include the Galapagos Islands and Bermuda. In a sense it is
442 reassuring that the SDMs are failing to pick up these areas, because they correspond to
443 regions with known adverse conditions for present-day formation of reefs and currently
444 support low coral biodiversity (Glynn *et al.*, 1983; Logan & Tomascik, 1991).
445
446 With a large enough number of explanatory variables, a statistical model may succeed in
447 explaining data by 'finding' causal relationships that are in fact dubious or nonexistent.
448 The process is known as over-fitting, and weakens the utility and applicability of the
449 model. Environmental variables having an effect which appear in conflict with biological
450 knowledge of coral habitat and physiology may indicate this occurring and hence can be
451 identified (and omitted) on this basis. This appears to be the case with the variables
452 included as indicators of upwelling (nutrients at 100m depth and dissolved oxygen both
453 at the surface and at 100m). By comparing the prediction maps for the *SST* and the
454 *SST_upwllng* models (Figs. 6a & c and 7a & c) it is obvious that the addition of the
455 upwelling indicators give an improved model 'prediction'. The prediction of habitat
456 suitability for coral reefs improves most, however, in areas that are either distant to
457 upwelling processes (such as Bermuda or Northern Red Sea) or where upwelling

458 conditions are expected to pose a serious challenge to reef presence, rather than be an
459 asset (e.g. Galapagos Islands). We highlight this example as a warning of how overfitting
460 could obscure important variables and responses in SDM. As a result we have opted for
461 not including these variables in our final predictor set and 'optimal' model (*OPTIMAL*).
462
463 **Geographical bias and error analysis of mapped SDM predictions**
464 Misclassification maps (Fig. 4; based on the false positives and false negatives averaged
465 from evaluation data of 30 model runs) were used to test for geographical bias in the
466 predicted probabilities. From visual inspection, no systematic bias is apparent between
467 major coral provinces (i.e. Caribbean and Indo-Pacific regions); any clustering of errors
468 takes place at lower spatial scales. All three SDM techniques, MaxEnt, CART and BRT,
469 fail to model suitable habitat for coral reefs at certain sites for which there are ReefBase
470 entries (false negatives). Interestingly, the percentage of false negatives is much higher
471 for cells that according to the reef location database correspond to non reef-forming
472 coral communities. For BRT (and MaxEnt), a total of 53% (65%) of cells with non-reef
473 coral communities give false negatives, compared to only 37% (45%) of the remaining
474 presence sites. Non-reef entries largely correspond to sites classified as 'marginal' by
475 Kleypas *et al*. (1999) and include the Pacific coast of Costa Rica and Panama, Bermuda,
476 the southern coasts of the main islands of Japan, southern Queensland, Australia and the
477 Galápagos Islands. Reef formation at these sites is known to be challenging (e.g.
478 mainland Japan, Veron & Minchin, 1992; Solitary Islands, Harriott *et al.*, 1994;
479 Galápagos, Glynn *et al.*, 1983), and typically coral communities are present rather than
480 reefs (ReefBase, 2000) because of an inability to accumulate $CaCO_3$ and a diminished

481 presence of primary reef-building taxa such as *Porites* and *Acropora* (Buddemeier &
482 Smith, 1999; Harriott & Banks, 2002). On the Pacific coast of Costa Rica and Panama,
483 reef development is sporadic and mostly takes place around offshore islands. D'Croz &
484 Robertson (1997) suggest this might be due to temperature extremes associated to El
485 Niño events and frequent cool upwelling episodes, with the formation of reef in the
486 coast further prevented by high concentration of nutrients, likely caused by terrestrial
487 runoff. The consistent SDM results across models hint that this region is currently
488 unsuitable for coral reef accretion. Missing input variable(s) may also explain some
489 consistent SDM output errors. For example, all three methods misclassify cells to give
490 false negatives across the region of the West Timor Sea and Rowley Shoals, off NW
491 Australia, where coral reefs are known to be present (Spalding *et al.*, 2001).
492
493 **Performance comparison of the different SDM techniques**
494 All three models were able to produce distributions that are consistent with the major
495 regions of coral reef distribution. However, the three SDM techniques used in this paper
496 are static, and therefore our underlying assumption is that surface water coral reefs are
497 in equilibrium (or quasi-equilibrium) with their environment (e.g. Peterson *et al.*, 1999;
498 Pearman *et al.*, 2008). Non-equilibrium situations are often considered more realistic in
499 an ecological context (Guisan & Zimmermann, 2000; Austin, 2002), but the required
500 global environmental datasets do not exist with the temporal frequency required to run
501 dynamic simulation models (Botkin *et al.*, 2007). Instead, we considered transient
502 effects in the analysis by including them as additional predictive variables (e.g. tropical

503 storm frequency and intensity, standard deviations in SST at inter-annual and weekly
504 time scales). None were found to have a significant impact on the models.
505
506 Presence-only methods, such as MaxEnt, provide a test of the potential anomalies in
507 model results created by inaccuracies or incompleteness of the input datasets. For
508 example, SDM results for wide-ranging and tolerant species can be particularly sensitive
509 to absence data (Brotons *et al.*, 2004). Hirzel *et al.* (2001) also showed that presence-
510 only models seemed to perform better than presence/absence ones in out-of-
511 equilibrium situations (i.e. when a species does not occupy all suitable habitat) and so
512 such models might be better suited to predict response to environmental change. A
513 limitation of static modelling is that the realised habitat is assumed to equal the
514 'fundamental' one (i.e. that coral reefs are present in all areas with an environment
515 compatible with requirements). However, even under equilibrium conditions, absence
516 might be due to dispersal limitations. For presence/absence models this can give rise to
517 an *a-priori* bias in the training data (Hutchinson, 1957; Pulliam, 2000; Phillips *et al.*,
518 2006), and in out-of-equilibrium situations migration limitations can become a
519 significant source of uncertainty that is also difficult to estimate (reviewed in Thuiller *et*
520 *al.*, 2008). Comparison with the output from presence-only methods can pinpoint the
521 magnitude of some of these effects on model performance (Gu & Swihart, 2004).
522 Encouragingly for our study, the presence-only method, MaxEnt, and the
523 presence/absence method, BRT, perform almost equally well (Table 2) in evaluation
524 tests (e.g. for ability to separate presence/absence signal) and in comparisons between
525 modelled predictions and actual presence/absence data. The application of both SDM

526 approaches in parallel, therefore, increases confidence in the models as predictive tools
527 for understanding the global distribution of surface water coral reefs under present
528 conditions.

529

530 ## Conclusions

531 We have used three SDM methods to predict coral reef presence/absence based on
532 environmental factors at 1° resolution. At this scale, SST-related variables are found to
533 be dominant in explaining the current biogeography of coral reefs, but alone are
534 insufficient to give an accurate picture of global distribution. Surface nutrient
535 concentration, aragonite saturation and light availability prove critical in accurately
536 defining the limits of coral reef ecosystems in both the Caribbean and Pacific. We find
537 that light availability is particularly crucial for the South Pacific, with the exception of
538 the Southern GBR, where nutrient concentrations are key. The Coral Triangle region is
539 poorly modelled using just SST variables, and an accurate prediction requires a
540 combination of nutrients, light availability and most significantly, aragonite saturation.

541

542 We have tested model performance on evaluation data, and find that BRT, a
543 presence/absence technique, performs slightly better than presence-only MaxEnt. Our
544 prediction confidence is strengthened by this result because any major inconsistencies
545 between the two would indicate issues associated with the input data, underlying
546 assumptions, or model selection.

547

548 Although the 1°x1° scale of our analysis is too coarse to capture local environmental
549 controls at reef sites and their impact on local-scale distribution, our chosen resolution
550 is consistent with the current generation of global climate models (ca. 1°x1°) and the
551 available global environmental data sets. Aggregation to coarser spatial scales also
552 increases the reliability of a species absence (McPherson *et al.*, 2006), and can thus
553 offset limitations in the reef location database. Explicit linking of environmental-
554 biogeographical and climate models will facilitate the generation of future (and past)
555 projections (and sensitivities) of coral reef biogeography to global environmental
556 change.
557

558 ## References

559 Antonov, J.I., Locarnini, R.A., Boyer, T.P., Mishonov, A.V. & Garcia, H.E. (2006) *World*
560 *Ocean Atlas 2005, Volume 2: Salinity*, (ed. by S. Levitus) edn. NOAA Atlas NESDIS,
561 U.S. Government Printing Office, Washington D.C.
562 Becker, J.J., Sandwell, D.T., Smith, W.H.F., Braud, J., Binder, B., Depner, J., Fabre, D.,
563 Factor, J., Ingalls, S., Kim, S.H., Ladner, R., Marks, K., Nelson, S., Pharaoh, A.,
564 Trimmer, R., Von Rosenberg, J., Wallace, G. & Weatherall, P. (2009) Global
565 Bathymetry and Elevation Data at 30 Arc Seconds Resolution: SRTM30_PLUS.
566 *Marine Geodesy*, **32**, 355-371.
567 Berkelmans, R., De'ath, G., Kininmonth, S. & Skirving, W.J. (2004) A comparison of the
568 1998 and 2002 coral bleaching events on the Great Barrier Reef: spatial
569 correlation, patterns, and predictions. *Coral Reefs*, **23**, 74-83.

570 Bindoff, N.L., Willebrand, J., V. Artale, Cazenave, A., Gregory, J., Gulev, S., Hanawa, K.,
571 Quéré, C.L., Levitus, S., Nojiri, Y., Shum, C.K., Talley, L.D. & Unnikrishnan, A.
572 (2007) Observations: Oceanic Climate Change and Sea Level. *Climate Change*
573 *2007: The Physical Science Basis* (ed. by S. Solomon, et al.). Cambridge University
574 Press, Cambridge, United Kingdom and New York, NY, USA.
575 Bishop, J.K.B. & Rossow, W.B. (1991) Spatial and temporal variability of global surface
576 solar irradiance. *Journal of Geophysical Research-Oceans*, **96**, 16839-16858.
577 Bishop, J.K.B., Rossow, W.B. & Dutton, E.G. (1997) Surface solar irradiance from the
578 International Satellite Cloud Climatology Project 1983-1991. *Journal of*
579 *Geophysical Research-Atmospheres*, **102**, 6883-6910.
580 Bivand, R.S., Pebesma, E.J. & Gomez-Rubio, V. (2008) *Applied Spatial Data Analysis with*
581 *R*. Springer, New York.
582 Botkin, D.B., Saxe, H., Araújo, M.B., Betts, R., Bradshaw, R.H.W., Cedhagen, T., Chesson, P.,
583 Dawson, T.P., Etterson, J.R., Faith, D.P., Ferrier, S., Guisan, A., Hansen, A.S., Hilbert,
584 D.W., Loehle, C., Margules, C., New, M., Sobel, M.J. & Stockwell, D.R.B. (2007)
585 Forecasting the effects of global warming on biodiversity. *Bioscience*, **57**, 227-
586 236.
587 Bradley, A.P. (1997) The use of the area under the roc curve in the evaluation of
588 machine learning algorithms. *Pattern Recognition*, **30**, 1145-1159.
589 Bradley, B.A. (2010) Assessing ecosystem threats from global and regional change:
590 hierarchical modeling of risk to sagebrush ecosystems from climate change, land
591 use and invasive species in Nevada, USA. *Ecography*, **33**, 198-208.

592 Breiman, L., Friedman, J., Stone, C. & Olshen, R.A. (1984) *Classification and Regression*
593 *Trees*. Wadsworth & Brooks/Cole Advanced Books & Software, Monterey,
594 California.
595 Brotons, L., Thuiller, W., Araújo, M.B. & Hirzel, A.H. (2004) Presence-absence versus
596 presence-only modelling methods for predicting bird habitat suitability.
597 *Ecography*, **27**, 437-448.
598 Cao, L. & Caldeira, K. (2008) Atmospheric CO_2 stabilization and ocean acidification.
599 *Geophysical Research Letters*, **35**, L19609.
600 Carpenter, K.E., Abrar, M., Aeby, G., Aronson, R.B., Banks, S., Bruckner, A., Chiriboga, A.,
601 Cortes, J., Delbeek, J.C., DeVantier, L., Edgar, G.J., Edwards, A.J., Fenner, D.,
602 Guzman, H.M., Hoeksema, B.W., Hodgson, G., Johan, O., Licuanan, W.Y.,
603 Livingstone, S.R., Lovell, E.R., Moore, J.A., Obura, D.O., Ochavillo, D., Polidoro, B.A.,
604 Precht, W.F., Quibilan, M.C., Reboton, C., Richards, Z.T., Rogers, A.D., Sanciangco,
605 J., Sheppard, A., Sheppard, C., Smith, J., Stuart, S., Turak, E., Veron, J.E.N., Wallace,
606 C., Weil, E. & Wood, E. (2008) One-third of reef-building corals face elevated
607 extinction risk from climate change and local impacts. *Science*, **321**, 560-563.
608 Carton, J.A. & Giese, B.S. (2008) A reanalysis of ocean climate using Simple Ocean Data
609 Assimilation (SODA). *Monthly Weather Review*, **136**, 2999-3017.
610 Cooper, T.F., De 'Ath, G., Fabricius, K.E. & Lough, J.M. (2008) Declining coral calcification
611 in massive Porites in two nearshore regions of the northern Great Barrier Reef.
612 *Global Change Biology*, **14**, 529-538.

613 Costa, G.C., Nogueira, C., Machado, R.B. & Colli, G.R. (2010) Sampling bias and the use of
614 ecological niche modeling in conservation planning: a field evaluation in a
615 biodiversity hotspot. *Biodiversity and Conservation*, **19**, 883-899.
616 D'Croz, L. & Robertson, D.R. (1997) Coastal oceanographic conditions affecting coral
617 reefs on both sides of the isthmus of Panama. *8th International Coral Reef*
618 *Symposium* (ed by H.A. Lessios & I.G. Macintyre), pp. 2053-2058.
619 Dana, F.D. (1843) On the temperature limiting the distribution of corals. *American*
620 *Journal of Science*, **45**, 130-131.
621 De'ath, G. (2007) Boosted trees for ecological modeling and prediction. *Ecology*, **88**,
622 243-251.
623 De'ath, G. & Fabricius, K.E. (2000) Classification and regression trees: A powerful yet
624 simple technique for ecological data analysis. *Ecology*, **81**, 3178-3192.
625 De'ath, G. & Fabricius, K. (2010) Water quality as a regional driver of coral biodiversity
626 and macroalgae on the Great Barrier Reef. *Ecological Applications*, **20**, 840-850.
627 De'ath, G., Lough, J.M. & Fabricius, K.E. (2009) Declining Coral Calcification on the Great
628 Barrier Reef. *Science*, **323**, 116-119.
629 Dickson, A.G. & Millero, F.J. (1987) A comparison of the equilibrium constants for the
630 dissociation of carbonic acid in seawater media. *Deep-Sea Research Part a -*
631 *Oceanographic Research Papers*, **34**, 1733-1743.
632 Elith, J., Leathwick, J.R. & Hastie, T. (2008) A working guide to boosted regression trees.
633 *Journal of Animal Ecology*, **77**, 802-813.
634 Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J.,
635 Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A.,

636 Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T.,
637 Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberon, J.,
638 Williams, S., Wisz, M.S. & Zimmermann, N.E. (2006) Novel methods improve
639 prediction of species' distributions from occurrence data. *Ecography*, **29**, 129-
640 151.
641 Ficetola, G.F., Maiorano, L., Falcucci, A., Dendoncker, N., Boitani, L., Padoa-Schioppa, E.,
642 Miaud, C. & Thuiller, W. (2010) Knowing the past to predict the future: land-use
643 change and the distribution of invasive bullfrogs. *Global Change Biology*, **16**, 528-
644 537.
645 Fielding, A.H. & Bell, J.F. (1997) A review of methods for the assessment of prediction
646 errors in conservation presence/absence models. *Environmental Conservation*,
647 **24**, 38-49.
648 Fitt, W.K., Brown, B.E., Warner, M.E. & Dunne, R.P. (2001) Coral bleaching:
649 interpretation of thermal tolerance limits and thermal thresholds in tropical
650 corals. *Coral Reefs*, **20**, 51-65.
651 Franklin, J. (2009) *Mapping Species Distribution: Spatial Inference and Prediction*.
652 Cambridge University Press, New York.
653 Friedman, J.H. (2001) Greedy function approximation: A gradient boosting machine.
654 *Annals of Statistics*, **29**, 1189-1232.
655 Friedman, J.H. & Meulman, J.J. (2003) Multiple additive regression trees with application
656 in epidemiology. *Statistics in Medicine*, **22**, 1365-1381.

657 Garcia, H.E., Locarnini, R.A., Boyer, T.P. & Antonov, J.I. (2006a) *World Ocean Atlas 2005,*
658 *Volume 4: Nutrients (phosphate, nitrate, silicate)*, (ed. by S. Levitus) edn. NOAA
659 Atlas NESDIS 64, U.S. Government Printing Office.
660 Garcia, H.E., Locarnini, R.A., Boyer, T.P. & Antonov, J.I. (2006b) *World Ocean Atlas 2005,*
661 *Volume 3: Dissolved Oxygen, Apparent Oxygen Utilization, and Oxygen Saturation*,
662 (ed. by S. Levitus) edn. NOAA Atlas NESDIS 63, U.S. Government Printing Office.
663 GlobColour (2008) Diffuse attenuation coefficient at 490 nm (KD490). In. Observation
664 de la Terre - Environnement (ACRI-ST), Sophia Antipolis
665 Glynn, P.W. & Colgan, M.W. (1992) Sporadic disturbances in fluctuating coral reef
666 environments: El Niño and coral reef development in the Eastern Pacific
667 *American Zoologist*, **32**, 707-718.
668 Glynn, P.W., Wellington, G.M. & Wells, J.W. (1983) *Corals and coral reefs of the Galápagos*
669 *Islands*. University of California Press.
670 Gu, W. & Swihart, R.K. (2004) Absent or undetected? Effects of non-detection of species
671 occurrence on wildlife-habitat models. *Biological Conservation*, **116**, 195-203.
672 Guisan, A. & Zimmermann, N.E. (2000) Predictive habitat distribution models in
673 ecology. *Ecological Modelling*, **135**, 147-186.
674 Hatcher, B.G. (1988) Coral reef primary productivity: A beggar's banquet. *Trends in*
675 *Ecology & Evolution*, **3**, 106-111.
676 Hernandez, P.A., Graham, C.H., Master, L.L. & Albert, D.L. (2006) The effect of sample
677 size and species characteristics on performance of different species distribution
678 modeling methods. *Ecography*, **29**, 773-785.

679 Hoegh-Guldberg, O., Mumby, P.J., Hooten, A.J., Steneck, R.S., Greenfield, P., Gomez, E.,
680 Harvell, C.D., Sale, P.F., Edwards, A.J., Caldeira, K., Knowlton, N., Eakin, C.M.,
681 Iglesias-Prieto, R., Muthiga, N., Bradbury, R.H., Dubi, A. & Hatziolos, M.E. (2007)
682 Coral reefs under rapid climate change and ocean acidification. *Science*, **318**,
683 1737-1742.
684 Johannes, R.E., Wiebe, W.J., Crossland, C.J., Rimmer, D.W. & Smith, S.V. (1983) Latitudinal
685 limits of coral reef growth. *Marine Ecology-Progress Series*, **11**, 105-111.
686 Johnson, C.J. & Gillingham, M.P. (2005) An evaluation of mapped species distribution
687 models used for conservation planning. *Environmental Conservation*, **32**, 117-
688 128.
689 Key, R.M., Kozyr, A., Sabine, C.L., Lee, K., Wanninkhof, R., Bullister, J.L., Feely, R.A.,
690 Millero, F.J., Mordy, C. & Peng, T.H. (2004) A global ocean carbon climatology:
691 Results from Global Data Analysis Project (GLODAP). *Global Biogeochemical*
692 *Cycles*, **18**, GB4031.
693 Kleypas, J.A. (2007) Constraints on Predicting Coral Reef Response to Climate Change.
694 *Geological Approaches to Coral Reef Ecology* (ed. by R. Aronson), pp. 386-424.
695 Springer, New York.
696 Kleypas, J.A., McManus, J.W. & Menez, L.A.B. (1999) Environmental limits to coral reef
697 development: Where do we draw the line? *American Zoologist*, **39**, 146-159.
698 Kleypas, J.A., Feely, R.A., Fabry, V.J., Langdon, C., Sabine, C.L. & Robbins, L.L. (2006)
699 Impacts of Ocean Acidification on Coral Reefs and Other Marine Calcifiers. A
700 Guide for Future Research. Report of a workshop sponsored by NSF, NOAA &
701 USGS.

Kuffner, I.B., Andersson, A.J., Jokiel, P.L., Rodgers, K.S. & Mackenzie, F.T. (2008) Decreased abundance of crustose coralline algae due to ocean acidification. *Nature Geoscience*, **1**, 114-117.

Leathwick, J.R., Elith, J., Francis, M.P., Hastie, T. & Taylor, P. (2006) Variation in demersal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees. **321**, 267-281.

Locarnini, R.A., Mishonov, A.V., Antonov, J.I., Boyer, T.P. & Garcia, H.E. (2006) *World Ocean Atlas 2005, Volume 1: Temperature*, (ed. by S. Levitus) edn. NOAA Atlas NESDIS 61, U.S. Government Printing Office, Washington, D.C.

Logan, A. & Tomascik, T. (1991) Extension Growth-Rates in two Coral Species from High-Latitude Reefs of Bermuda. *Coral Reefs*, **10**, 155-160.

Mahowald, N.M., Muhs, D.R., Levis, S., Rasch, P.J., Yoshioka, M., Zender, C.S. & Luo, C. (2006) Change in atmospheric mineral aerosols in response to climate: Last glacial period, preindustrial, modern, and doubled carbon dioxide climates. *Journal of Geophysical Research-Atmospheres*, **111**, D10202.

Manel, S., Williams, H.C. & Ormerod, S.J. (2001) Evaluating presence-absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology*, **38**, 921-931.

McCook, L.J., Jompa, J. & Diaz-Pulido, G. (2001) Competition between corals and algae on coral reefs: a review of evidence and mechanisms. *Coral Reefs*, **19**, 400-417.

Mehrbach, C., Culberso.Ch, Hawley, J.E. & Pytkowic.Rm (1973) Measurement of apparent dissociation-constants of carbonic-acid in seawater at atmospheric-pressure. *Limnology and Oceanography*, **18**, 897-907.

725 Moberg, F. & Folke, C. (1999) Ecological goods and services of coral reef ecosystems.
726 *Ecological Economics*, **29**, 215-233.
727 Moisen, G.G., Freeman, E.A., Blackard, J.A., Frescino, T.S., Zimmermann, N.E. & Edwards,
728 T.C. (2006) Predicting tree species presence and basal area in Utah: A
729 comparison of stochastic gradient boosting, generalized additive models, and
730 tree-based methods. *Ecological Modelling*, **199**, 176-187.
731 Pebesma, E.J. & Bivand, R.S. (2005) Classes and methods for spatial data in R. *R News*, **5**,
732 9-13.
733 Phillips, S.J. (2006) A brief tutorial on Maxent. AT & T Research. Available at:
734 www.cs.princeton.edu/~schapire/maxent/tutorial/tutorial.doc
735 Phillips, S.J. & Dudik, M. (2008) Modeling of species distributions with MaxEnt: new
736 extensions and a comprehensive evaluation. *Ecography*, **31**, 161-175.
737 Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2006) Maximum entropy modeling of
738 species geographic distributions. *Ecological Modelling*, **190**, 231-259.
739 R Development Core Team (2009) *R: A Language and Environment for Statistical*
740 *Computing*. R Foundation for Statistical Computing, Vienna, Austria.
741 Rayner, N.A., Brohan, P., Parker, D.E., Folland, C.K., Kennedy, J.J., Vanicek, M., Ansell, T.J.
742 & Tett, S.F.B. (2006) Improved analyses of changes and uncertainties in sea
743 surface temperature measured in situ sice the mid-nineteenth century: The
744 HadSST2 dataset. *Journal of Climate*, **19**, 446-469.
745 Reaka-Kudla, M.L. (1997) The global biodiversity of coral reefs: A comparison with rain
746 forests. *Biodiversity, II. Understanding and protecting our biological resources*, 83-
747 108.

748 ReefBase (2000) *ReefBase 2000: improving policies for sustainable management of coral*
749 *reefs*. International Centre for Living Aquatic Resources Management,
750 Philippines.
751 Reynolds, R.W., Rayner, N.A., Smith, T.M., Stokes, D.C. & Wang, W.Q. (2002) An improved
752 in situ and satellite SST analysis for climate. *Journal of Climate*, **15**, 1609-1625.
753 Ridgeway, G. (2007) *gbm: Generalized Boosted Regression Models*. R package version 1.6-
754 3.1.
755 Riegl, B. (2001) Inhibition of reef framework by frequent disturbance: examples from
756 the Arabian Gulf, South Africa, and the Cayman Islands. *Palaeogeography*
757 *Palaeoclimatology Palaeoecology*, **175**, 79-101.
758 Seo, C., Thorne, J.H., Hannah, L. & Thuiller, W. (2009) Scale effects in species distribution
759 models: implications for conservation planning under climate change. *Biology*
760 *Letters*, **5**, 39-43.
761 Shinn, E.A., Smith, G.W., Prospero, J.M., Betzer, P., Hayes, M.L., Garrison, V. & Barber, R.T.
762 (2000) African dust and the demise of Caribbean coral reefs. *Geophysical*
763 *Research Letters*, **27**, 3029-3032.
764 Smith, T.B. (2008) Temperature effects on herbivory for an Indo-Pacific parrotfish in
765 Panama: implications for coral-algal competition. *Coral Reefs*, **27**, 397-405.
766 Spalding, M.D. & Grenfell, A.M. (1997) New estimates of global and regional coral reef
767 areas. *Coral Reefs*, **16**, 225-230.
768 Spalding, M.D., Ravilious, C. & Green, E.P. (2001) *World atlas of coral reefs*, University of
769 California Press, Berkeley.

770 Swets, J.A. (1988) Measuring the Accuracy of Diagnostic Systems. *Science*, **240**, 1285-
771 1293.
772 Therneau, T.M. & Atkinson, B. (2010) *rpart: Recursive Partitioning*. R port by Brian
773 Ripley, R package version 3.1-48.
774 Thuiller, W. (2004) Patterns and uncertainties of species' range shifts under climate
775 change. *Global Change Biology*, **10**, 2020-2027.
776 Tittensor, D.P., Baco, A.R., Brewin, P.E., Clark, M.R., Consalvey, M., Hall-Spencer, J.,
777 Rowden, A.A., Schlacher, T., Stocks, K.I. & Rogers, A.D. (2009) Predicting global
778 habitat suitability for stony corals on seamounts. *Journal of Biogeography*, **36**,
779 1111-1128.
780 Turley, C., Eby, M., Ridgwell, A.J., Schmidt, D.N., Findlay, H.S., Brownlee, C., Riebesell, U.,
781 Fabry, V.J., Feely, R.A. & Gattuso, J.P. (2010) The societal challenge of ocean
782 acidification. *Marine Pollution Bulletin*, **60**, 787-792.
783 UNEP/DEWA/GRID-Europe (2009) Tropical cyclones average sum of windspeed 1975-
784 2007, in *Global Assessment Report on Disaster Risk Reduction, Appendix 1*, United
785 Nations, Geneva, Switzerland.
786 van Heuven, S., Pierrot, D., Lewis, E. & Wallace, D.W.R. (2009) *MATLAB Program*
787 *Developed for CO_2 System Calculations*. Carbon Dioxide Information Analysis
788 Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge,
789 Tennessee.
790 Weaver, A.J., Eby, M., Wiebe, E.C., Bitz, C.M., Duffy, P.B., Ewen, T.L., Fanning, A.F., Holland,
791 M.M., MacFadyen, A., Matthews, H.D., Meissner, K.J., Saenko, O., Schmittner, A.,
792 Wang, H.X. & Yoshimori, M. (2001) The UVic Earth System Climate Model: Model

description, climatology, and applications to past, present and future climates. *Atmosphere-Ocean*, **39**, 361-428.

Zweig, M.H. & Campbell, G. (1993) Receiver-Operating Characteristic (Roc) Plots - a Fundamental Evaluation Tool in Clinical Medicine. *Clinical Chemistry*, **39**, 561-577.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1 BRT model optimisation, output and analysis.

Appendix S2 MaxEnt model output and analysis.

Appendix S3 Environmental variable correlations and contributions.

Acknowledgements

E.C. is funded by a U. Bristol PhD scholarship, A.R. by the Royal Society, and E.J.H by a RCUK Academic Fellowship and additional support from a NERC Fellowship (NE/D010012/1). The authors greatly valued the constructive insights and suggestions of the Reviewers and the Editor. E.C. also wishes to acknowledge Hannah Rose and Dr. Kelvyn Jones for their suggestions and help regarding SDM methods.

816

817 Biosketch

818 **Elena Couce**

819 Elena Couce is a joint PhD candidate between the Department of Earth Sciences and
820 School of Geographical Sciences at the University of Bristol. Her research focuses on
821 studying the environmental requirements for coral reef ecosystems, and modelling the
822 short-term impacts of climate change.

823

824 **Andy Ridgwell**

825 Andy Ridgwell is Professor in Earth System Modelling in the School of Geographical
826 Sciences, Bristol University. His main research interest is in accessing the causes and
827 consequences of changes in global carbon cycling and the role of organisms and
828 ecosystems in this.

829

830 **Erica J. Hendy**

831 Erica Hendy has a joint RCUK academic fellowship between the School of Biological
832 Sciences and the Department of Earth Sciences, University of Bristol. Her research
833 interest in coral reefs span from probing coral calcification using novel isotopic tools to
834 global-scale responses to climate change.

835

836 Author contributions: E.C., A.R. and E.J.H. conceived the research idea and collaborated
837 in the interpretation of the analysis, E.C. led the writing, and undertook the data
838 processing and modelling.

839 Editor: David Bellwood

840

Figure headings:

Figure 1. Mask used for the study (grey), corresponding to shallow waters between -60° and 60° latitude. Presence cells (those containing at least one entry from the ReefBase version 2000 dataset) are shown in black. The areas not included in the analysis are white. Behrmann Projection.

Figure 2. Predicted probability ((a) BRT and (c) CART) or logistic prediction ((b) MaxEnt) of a coral reef being present based on the environmental variables from the *OPTIMAL* model variable dataset (Table 1) and using the complete dataset for training. Each cell is designated a likelihood value for the predicted probability or logistic prediction from 0 (unlikely; in blue) to 1 (highly likely habitat; in red).

Figure 3. Histogram for the average modelled probabilities (BRT and CART), or logistic prediction (MaxEnt), of a coral reef being present in the shallow-water grid cells using BRT (a), MaxEnt (b) or CART (c). Cells where coral reefs or communities are present according to the ReefBase dataset are shown in grey, absence sites in white. The probability (logistic prediction in MaxEnt's case) assigned to each cell is the average of the runs in which it is taken as evaluation data out of a total of 30 random runs. The distribution of model results that are false positives (red) or false negatives (blue) is also shown.

862 Figure 4. Maps showing the spatial distribution of false positives (red) and false
863 negatives (blue) obtained for evaluation cells in 30 runs of BRT (a), MaxEnt (b) and
864 CART (c) models. Areas in grey in the MaxEnt map indicate cells where environmental
865 data was missing (e.g. areas missing aragonite saturation data) because predictions can
866 not be obtained using MaxEnt at sites with missing environmental data; in contrast
867 tree-based techniques are able to look for alternative variables.
868
869 Figure 5. ROC curves for the three model techniques and using 25% evaluation data
870 chosen randomly for each of 10 model runs. The optimal point in ROC space for BRT
871 (circles) and CART (triangles) corresponds to a threshold value of 50% probability of
872 presence.
873
874 Figure 6. Prediction maps for BRT models using various combinations of environmental
875 variables to evaluate predictor contributions to model performance. The predictor
876 subset used in each of the models are given in Table 1; SST variables (a), SST variables
877 plus aragonite saturation state (b), SST plus upwelling variables (c), SST plus light
878 variables (d), SST plus nutrient variables, (e) SST plus secondary SST variables
879 capturing SST variability on various timescales (f).
880
881 Figure 7. Prediction maps for MaxEnt models (given in Table 1) using various
882 combinations of environmental data to evaluate individual variable contributions to
883 model performance. As for Figure 6, the predictor subset used in each of the models are
884 given in Table 1; SST variables (a), SST variables plus aragonite statuation state (b),

885 SST plus upwelling variables (c), SST plus light variables (d), SST plus nutrient
886 variables, (e) SST plus secondary SST variables capturing SST variability on various
887 timescales (f).

Figure 1.

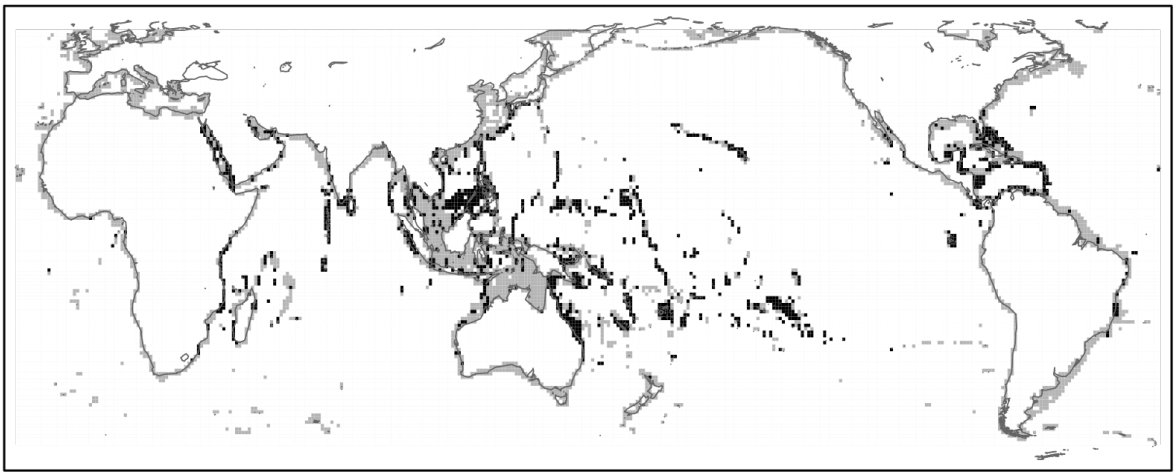


Figure 2.

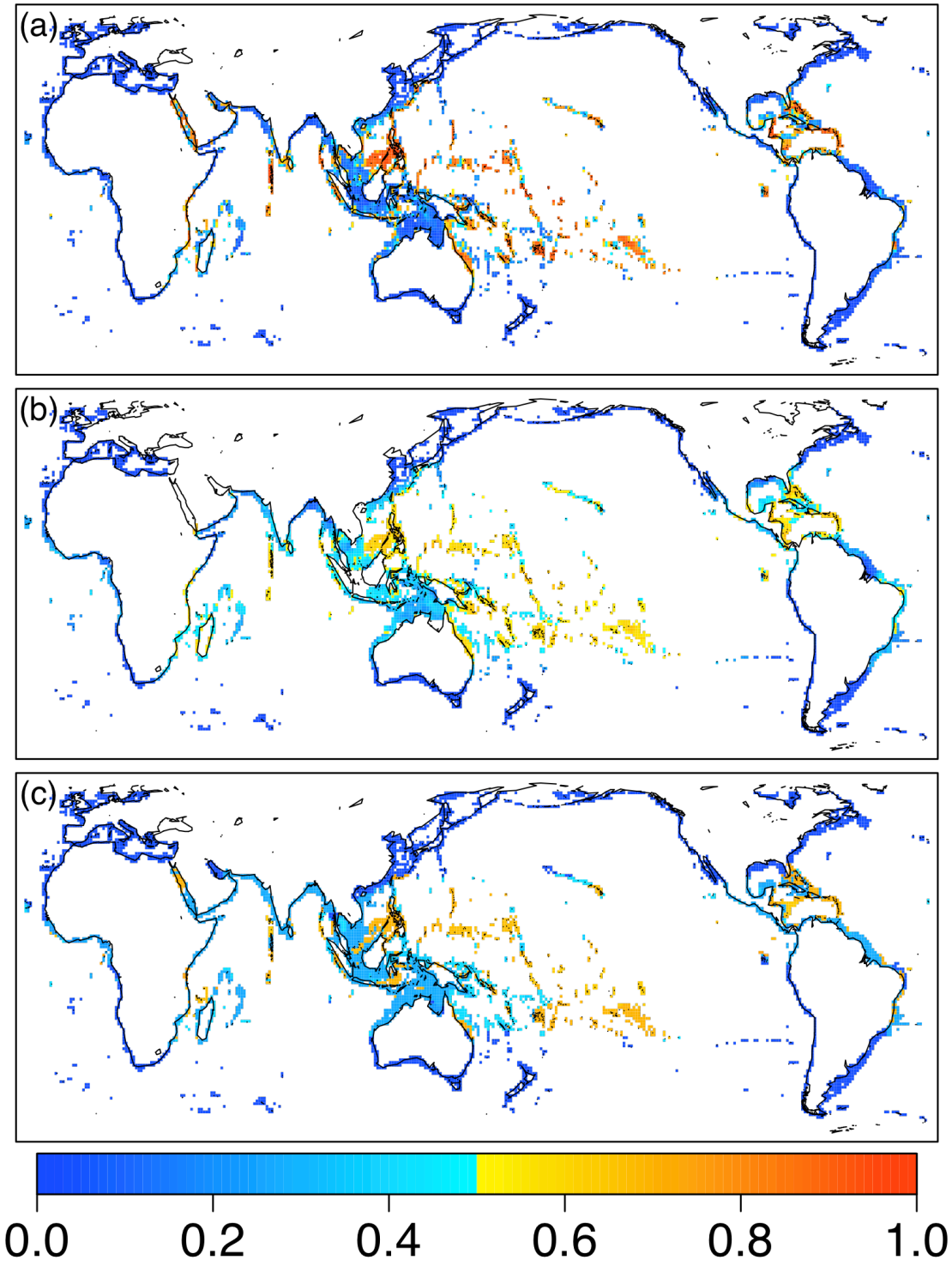


Figure 3.

-

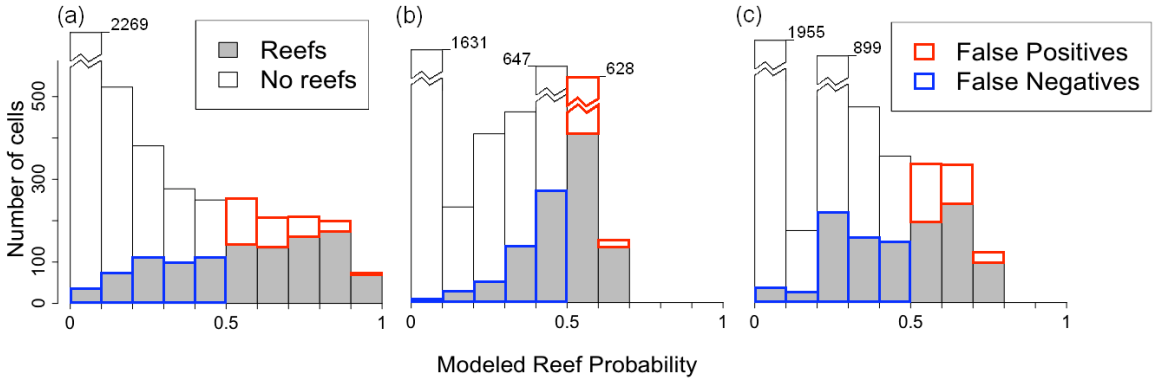

(a)
2269
(b)
1631
647
628
(c)
1955
899
Reefs
No reefs
False Positives
False Negatives
Number of cells
0
100
300
500
0
0.5
1
Modeled Reef Probability

Figure 4.

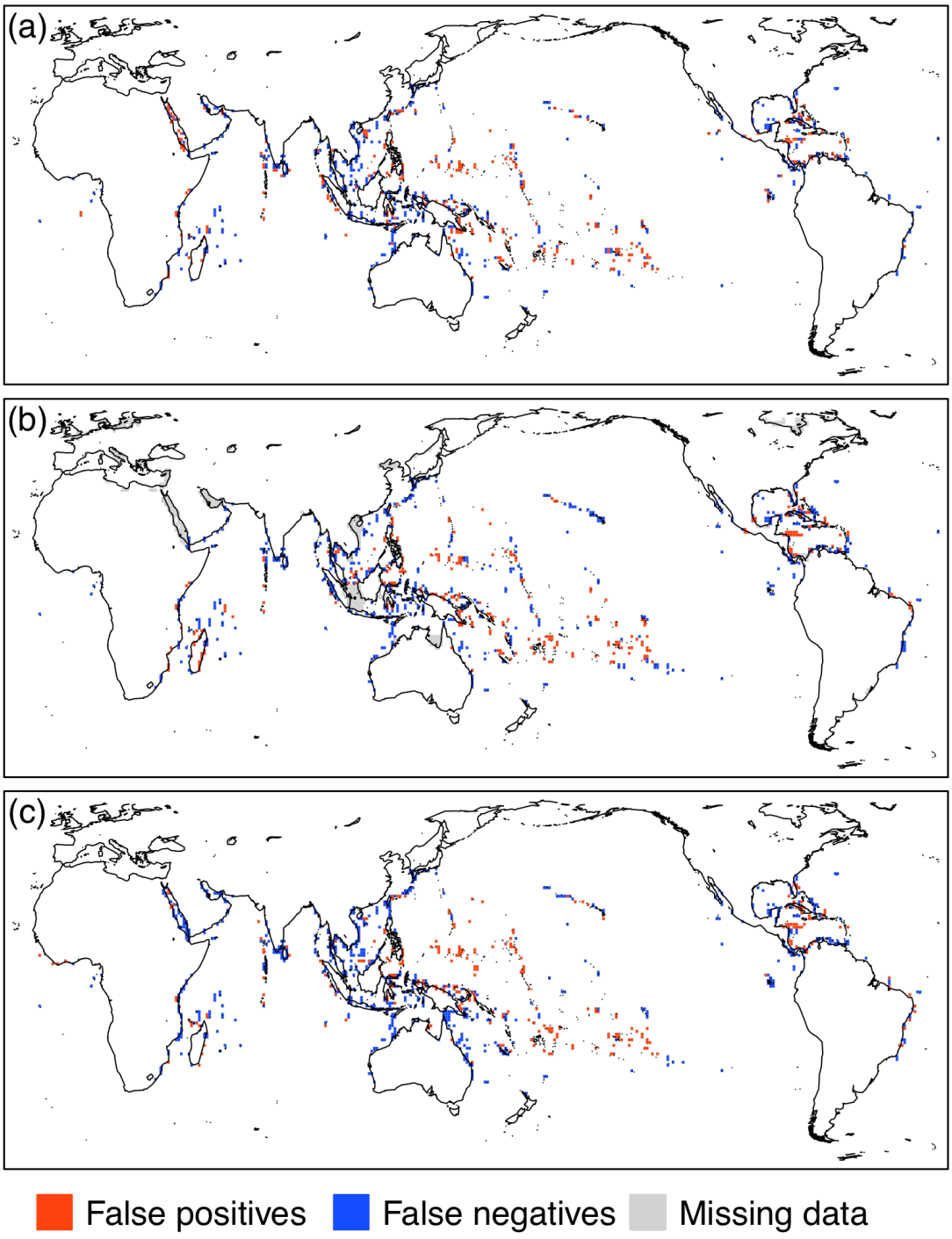
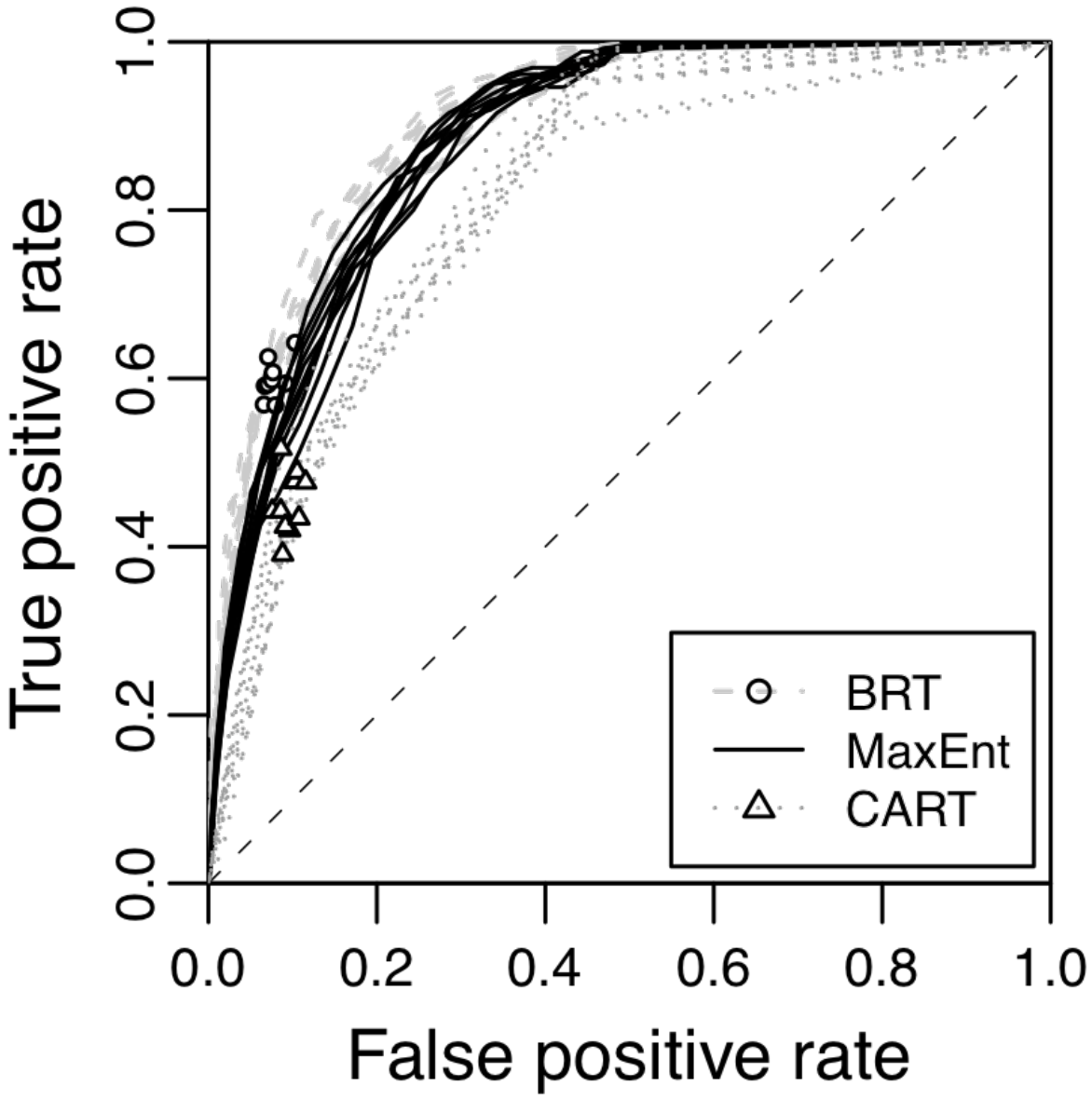


Figure 5.


True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0
BRT
MaxEnt
CART